

**IMPLEMENTASI HIRARKI DATASET DALAM  
MEMBANGUN MODEL LANGUAGE AKSARA  
BALI MENGGUNAKAN FRAMEWORK  
TESSERACT OCR**

**TESIS**



Oleh  
**AHMAD ASRONI**  
**NIM 2129101009**

**PROGRAM STUDI ILMU KOMPUTER  
PASCASARJANA  
UNIVERSITAS PENDIDIKAN GANESHA  
2024**

**IMPLEMENTASI HIRARKI DATASET DALAM  
MEMBANGUN MODEL LANGUAGE AKSARA  
BALI MENGGUNAKAN FRAMEWORK  
TESSERACT OCR**

**TESIS**

Diajukan kepada Universitas Pendidikan Ganesha  
untuk Memenuhi Sebagian Persyaratan  
Memperoleh Gelar Magister Komputer  
Program Studi Ilmu Komputer



Oleh  
**AHMAD ASRONI**  
**NIM 2129101009**

**PROGRAM STUDI ILMU KOMPUTER  
PASCASARJANA  
UNIVERSITAS PENDIDIKAN GANESHA  
2024**

## LEMBAR PERSETUJUAN PEMBIMBING

Tesis oleh Ahmad Asroni ini telah diperiksa dan disetujui untuk mengikuti Ujian Tesis

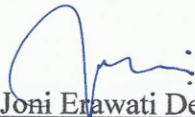
Singaraja, 31 Mei 2024

Pembimbing I,



Dr. Gede Indrawan., S.T., M. T.  
NIP 19760102 200312 1 001

Pembimbing II,




Dr. Luh Joni Erawati Dewi, S.T., M.Pd.  
NIP 19760625 200112 2 001

## LEMBAR PERSETUJUAN TIM PENGUJI

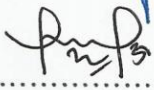
Tesis oleh Ahmad Asroni ini telah dipertahankan di depan tim penguji dan dinyatakan diterima sebagai salah satu persyaratan untuk memperoleh gelar Magister Komputer di Program Studi Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha,

Disetujui pada tanggal : Juli 2024

oleh  
Tim Penguji

.....  



Ketua (Dr. I Made Gede Sunarya, S.Kom, M.Cs.)  
NIP 198307252008011008

.....  


Anggota (Dr. I Gede Aris Gunadi, S.SI., M.Kom.)  
NIP 197703182008121004

.....  


Anggota (Dr. Gede Indrawan, S.T., M.T.)  
NIP 197601022003121001

.....  


Anggota (Dr. Luh Joni Erawati Dewi, S.T., M.PD.)  
NIP 197606252001122001

Mengetahui Direktur  
Program Pascasarjana Undiksha,  
  
Prof. Dr. I Nyoman Jampel, M.Pd.  
NIP 195910101986031003

## LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya bahwa tesis yang saya susun sebagai syarat untuk memperoleh gelar Magister Komputer dari Pascasarjana Universitas Pendidikan Ganesha seluruhnya merupakan hasil karya saya sendiri. Bagian-bagian tertentu dalam penulisan tesis yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas dan sesuai dengan norma, kaidah, serta etika akademis.

Apabila di kemudian hari ditemukan seluruh atau sebagian tesis ini bukan hasil karya saya sendiri atau adanya plagiat dalam bagian-bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sandang dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku di wilayah Negara Kesatuan Republik Indonesia.

Singaraja, 22 Juli 2024

Yang memberi pernyataan,



(Ahmad Asroni)



## PRAKATA

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas anugrah- Nya, sehingga tesis yang berjudul: “Implementasi Hirarki *Dataset* Dalam Membangun *Model Language* Aksara Bali Menggunakan *Framework Tesseract OCR*”, dapat diselesaikan sesuai dengan yang direncanakan.

Tesis ini ditulis untuk memenuhi salah satu persyaratan untuk memperoleh gelar Magister Komputer Pascasarjana Universitas Pendidikan Ganesha pada Program Studi Ilmu Komputer. terselesaikannya tesis ini telah banyak memperoleh uluran tangan dari berbagai pihak. Untuk itu, ijin penulis menyampaikan terima kasih dan penghargaan kepada pihak-pihak berikut.

1. Dr. Gede Indrawan, S.T., M.T., sebagai pembimbing I yang telah dengan sabar membimbing, mengarahkan, dan memberikan motivasi yang demikian bermakna, sehingga penulis mampu melewati berbagai hambatan dalam perjalanan studi dan penyelesaian tesis ini.
2. Dr. Luh Joni Erawati Dewi, S.T., M.Pd., sebagai pembimbing II, yang dengan gaya dan pola komunikasi yang khas, telah melecut semangat, motivasi, dan harapan penulis selama penelitian dan penulisan naskah laporan tesis ini, sehingga tesis ini dapat terwujud dengan baik sesuai harapan.
3. Dr. I Gede Aris Gunadi, S.SI., M.Kom. dan Dr. I Made Gede Sunarya, S.Kom, M.Cs. sebagai penguji yang telah banyak memberikan masukan-masukan yang bermanfaat untuk penyempurnaan tesis ini.
4. Kepala PT. Laksita Emy Saguna yang telah memberikan izin dan bantuan administratif sehingga penulis memperoleh kesempatan melanjutkan studi dan menyelesaikan tesis ini.
5. Koordinator Program Studi Ilmu Komputer dan staf dosen pengajar yang telah banyak membantu dan memotivasi penulis selama penyusunan tesis ini.
6. Direktur Pascasarjana Undiksha dan staf, yang telah banyak membantu selama penulis menyelesaikan tesis ini.
7. Rektor Universitas Pendidikan Ganesha, yang telah memberikan bantuan secara moral dan memfasilitasi berbagai kepentingan penulis dalam menyelesaikan tesis ini.

8. Rekan-rekan seangkatan di Program Studi Ilmu Komputer yang dengan karakternya masing-masing telah banyak berkontribusi membentuk kedirian penulis selama menjalani studi dan penyelesaian tesis ini.
9. Bapak Asmanan dan Ibu Zaitun selaku orang tua penulis, yang telah banyak membantu secara material dan moral selama penyelesaian tesis ini.
10. Rizki Nurmalia dan Isnawati selaku kakak penulis, yang telah banyak memberikan motivasi dan dorongan secara langsung maupun tidak langsung selama penyelesaian tesis ini.

Semoga semua bantuan yang telah mereka berikan dalam menyelesaikan studi ini, mereka diberkati imbalan yang sepadan oleh Tuhan Yang Maha Esa, kesehatan, dan keharmonian dalam menjalani kehidupan.

Penulis menyadari bahwa tesis ini belum sempurna. Namun, kehadirannya dalam konstelasi masyarakat akademis akan menambah perbendaharaan ilmu dalam perkembangan ilmu pengetahuan. Semoga tesis ini bermanfaat bagi masyarakat akademis, terutama mereka yang menyatakan diri bernaung di bawah kebesaran panji-panji pendidikan.

Singaraja, 22 Juli 2024

Penulis

## DAFTAR ISI

	<i>(halaman)</i>
LEMBAR PERSETUJUAN PEMBIMBING .....	i
LEMBAR PERSETUJUAN TIM PENGUJI.....	ii
LEMBAR PERNYATAAN .....	iii
PRAKATA.....	iv
ABSTRAK .....	vi
ABSTRACT.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR .....	x
DAFTAR TABEL.....	xi
DAFTAR LAMPIRAN.....	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Identifikasi Masalah.....	6
1.3 Batasan Masalah.....	7
1.4 Rumusan Masalah.....	7
1.5 Tujuan Penelitian.....	7
1.6 Manfaat Penelitian.....	8
1.7 Rencana Publikasi.....	8
BAB II KAJIAN PUSTAKA.....	10
2.1 Aksara Bali.....	10
2.2 <i>Unicode</i> .....	11
2.3 <i>Noto Sans Balinese</i> .....	11
2.4 <i>Web Scraping</i> .....	13
2.5 <i>Optical Character Recognition</i> .....	16
2.6 Tesseract OCR.....	20
2.7 Hierarki Dataset.....	22
2.8 Flutter Mobile Framework.....	23
2.9 Penelitian Relevan.....	24
BAB III METODE PENELITIAN.....	27
3.1 Dataset Preparation.....	28
3.2 Generate Dataset.....	29
3.3 Training Dataset.....	30

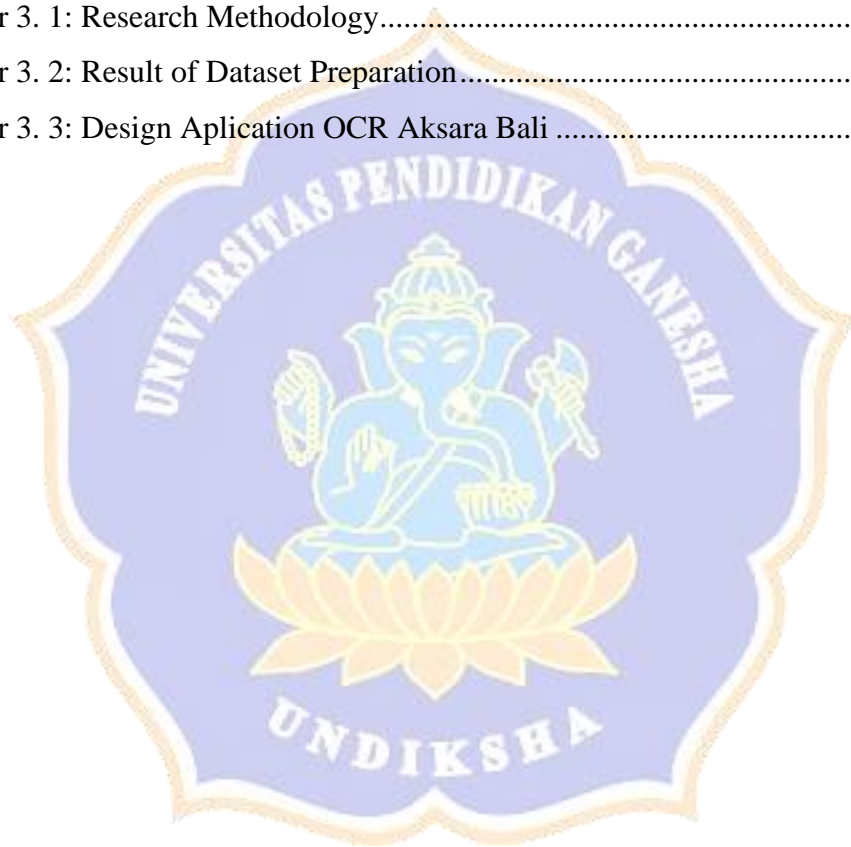


3.4	Testing.....	33
3.5	Implementation .....	34
BAB IV HASIL PENELITIAN .....		36
4.1	Generate Dataset .....	36
4.2	Training Dataset .....	45
4.3	Testing.....	49
4.4	Implementation .....	52
BAB V PENUTUP.....		54
5.1	Rangkuman .....	54
5.2	Simpulan .....	55
5.3	Saran.....	57
DAFTAR PUSTAKA .....		59
LAMPIRAN.....		65



## DAFTAR GAMBAR

Gambar 2. 1: Character Map Font Noto Sans Balinese .....	12
Gambar 2. 2: Unicode Noto Sans Balinese Font .....	13
Gambar 2. 3: Proses Umum <i>Optical Character Recognition</i> .....	17
Gambar 2. 4: Ilustrasi Alur Kerja Software OCR .....	20
Gambar 2. 5: Alur Proses Tesseract OCR.....	21
Gambar 2. 6: Arsitektur Dasar Flutter.....	24
Gambar 3. 1: Research Methodology.....	27
Gambar 3. 2: Result of Dataset Preparation.....	28
Gambar 3. 3: Design Aplication OCR Aksara Bali .....	35



## DAFTAR TABEL

Tabel 3. 1 Composition of Transliterated Dataset .....	30
--	----



## DAFTAR LAMPIRAN

Lampiran 1. Pameran Aplikasi Pada Acara DTIK Festival 2023 .....	65
Lampiran 2. Cuplikan Kode Web Scrapping .....	74

